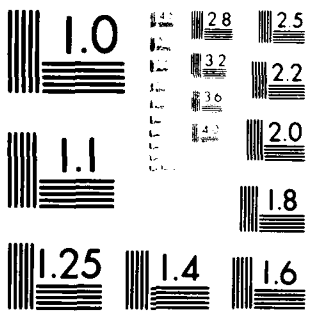
AD-A115 391 STANFORD UNIV CA DEPT OF STATISTICS F/G 12/1
UPDATING SUBJECTIVE PROBABILITY.(U)
MAR 82 P DIACONIS, S L ZABELL N00014-76-C-0475
UNCLASSIFIED TR-315 NL

END
DATE
FILMED
7 82
DTIC



MICROCOPY RESOLUTION TEST CHART

UPDATING SUBJECTIVE PROBABILITY

By

PERSI DIACONIS and SANDY L. ZABELL

TECHNICAL REPORT NO. 315

MARCH 2, 1982

Prepared Under Contract
N00014-76-C-0475 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

1. INTRODUCTION

1.1 Belief Revision

The most frequently discussed method of revising a subjective probability distribution P to obtain a new distribution P^*, based on the occurrence of an event E, is Bayes' rule: $P^*(A) = P(AE)/P(E)$. Richard Jeffrey (1965, 1968) has argued persuasively that Bayes' rule is not the only reasonable way to update: use of Bayes' rule presupposes that both $P(E)$ and $P(AE)$ have been previously quantified. In many instances this will clearly not be the case. Consider the following example:

Coin Tossing. Suppose we are thinking about three tosses of a coin. Under the usual circumstances a probability assignment is made on the eight possible outcomes $\Omega = \{000, 001, 010, 011, 100, 101, 110, 111\}$. Suppose an informant, believed trustworthy, announces: "Oh, I see you're thinking about that coin. I just spun it 100 times in the other room and it came up heads 80 times". This is clearly relevant information and we will obviously want to revise our opinion. The information cannot be put in terms of the occurrence of an event in the eight point space Ω and the Bayes rule is not directly available. Among many possible approaches, four methods of incorporating the information will be discussed.

a) Complete Reassessment.

b) Jeffrey's Rule.

c) Retrospective Conditioning.

d) Exchangeability.

a) Complete Reassessment. In the absence of further structure it is always possible to react to the new information by completely reassessing

P^* , presumably using the same techniques used to quantify the original distribution P.

b) Jeffrey's Rule. Suppose that the original probability assignment P was exchangeable. That is, $P(001) = P(010) = P(100)$ and $P(110) = P(101) = P(011)$. In the situation described, the information provided contains no information about the order of the next three tosses and thus we may well require the new probability distribution remain exchangeable. This is equivalent to considering a partition $\{E_i\}_{i=0}^3$ of Ω , where $E_0 = \{000\}$, $E_1 = \{001, 010, 100\}$, $E_2 = \{110, 101, 011\}$, $E_3 = \{111\}$. Here E_i is the set of outcomes with i ones and exchangeability implies that for any event A , and any i , $P(A|E_i) = P^*(A|E_i)$. To complete the probability assignment P^* , a subjective assessment of $P^*(E_i)$ is needed. Then P^* is determined by

$$P^*(A) = \Sigma P^*(A|E_i)\ P^*(E_i) = \Sigma P(A|E_i)\ P^*(E_i) \ .$$

The rule

$$P^*(A) = \Sigma P(A|E_i)\ P^*(E_i) \tag{1.1}$$

is known in the philosophical literature as Jeffrey's rule of conditioning. It is valid whenever there is a partition $\{E_i\}$ of the sample space such that

$$P^*(A|E_i) = P(A|E_i) \quad \text{for all } A \text{ and } i \ . \tag{J}$$

c) Retrospective Conditioning. Some subjectivists have suggested trying to analyze this kind of problem by momentarily disregarding the new information, quantifying a distribution on a space Ω^* rich enough to allow ordinary conditioning to be used, and then using Bayes' rule. For some discussion of this, see de Finetti (1972, Chap. 8) and Section 2.1 below. It is worth emphasizing that this type of retrospective conditioning is an extremely difficult psychological task; Fischoff (1975), and Fischoff and Beyth (1975) have demonstrated that "reporting the outcome of a historical event increases the perceived likelihood of that outcome", and Slovic and Fischoff (1977) have shown that "similar hindsight effects occur when people evaluate the predictability of scientific results -- they tend to believe they 'knew all along' what the experiments would find". Nor, in principle, is retrospective conditioning simpler than complete reassessment: since $P^*(A) = P(AE)/P(E)$ in this case, assessment of $P(AE)$ for each A is equivalent to reassessment of $P^*(A)$.

d) Exchangeability. The three future tosses of the coin may be regarded as exchangeable with the 100 tosses reported by our informant. Standard Bayesian computations can then be used.

Approaches b, c, and d are all special routes to the requantification of approach a; each is valid or useful under different assumptions. For example, Jeffrey's rule assumes the availability of a partition and the validity of assumption J. Retrospective conditioning assumes that one can do a reasonable job of assessing probabilities as if the data had not been observed. Exchangeability assumes that future trials are based on the same mechanism as past ones; in the example this might not be reasonable, perhaps the past trials were spins on a table, the future trials are tosses onto the floor.

In this paper we study the assumptions and conclusions that attend Jeffrey's rule. Our main contributions are technical: In Section 2 we connect Jeffrey's rule with sufficiency; Sections 3, 4, and 5 analyze what happens when two or more partitions are considered. In Section 3 we discuss commutativity of successive updating. In Section 4 we discuss methods for dealing with two partitions simultaneously, giving a necessary and sufficient condition for two probability measures on two algebras to have a common extension. In Section 5 we discuss some other motivations for Jeffrey's rule when condition (J) has not been subjectively checked. Jeffrey's rule gives the "closest" measure to P which fixes $P^*(E_i)$, and is related to the iterated proportional fitting procedure used in the statistical analysis of contingency tables. For ease of exposition, most of this paper assumes a countable state space or a countable partition $\{E_i\}_{i=1}^{\infty}$. In Section 6 we describe the mathematical machinery needed to extend the previous results to abstract probability spaces.

1.2 Historical and Bibliographical Note

We do not propose to survey here the growing philosophical literature on probability revision and Jeffrey's rule. The following quotations and references, however, should make clear that the problem was early recognized by the founders of modern subjective probability, and may be helpful as a guide to the recent literature.

From the subjectivistic perspective, the conditional probability $P(A|E)$ is the probability we currently would attribute to an event A if in addition to our present information we were also to learn E. In the language of betting, it is "the probability that we would regard as fair for a bet on A to be made immediately, but to become operative only if E occurs" (de Finetti, 1972, p. 193; cf. Ramsey 1931, p. 180). In this formulation,

the equality $P(A|E) = P(AE)/P(E)$ is not a definition, but follows as a theorem derived from the assumption of coherence (de Finetti, 1975, Chapter 4).

If we actually <u>learn</u> E to be true, it is conventional to adopt as one's new probability

$$P^*(A) = P(A|E) . \tag{1.2}$$

Assumption (1.2) seems entirely plausible - what else should our probability of A be, given that we have learned E , and nothing else, other than the probability which we were willing to attribute to A if we were subsequently to learn E? Several authors have pointed out that (1.2) is an <u>assumption</u>. Hacking (1967, p. 314) refers to (1.2) as the <u>dynamic assumption of personalism</u>, to contrast it with the static nature of the assumption of coherence. Hacking (1967, pp. 315-316) points out that coherence in its usual sense does not entail (1.2) and de Finetti concedes as much when he refers to an unexplained "criterion of temporal coherency" (de Finetti, 1972, p. 150); cf. Ramsey (1931, p. 192), who similarly asserts that "when my degrees of belief change in this way we can say that they have been changed consistently by my observation". For two attempts at a partial justification, however, see Freedman and Purves (1969), Teller (1976).

Ramsey himself perhaps stated the difficulty most clearly:

> [The degree of belief in <u>p</u> given <u>q</u>] is not the same as the degree to which [a subject] would believe <u>p</u>, if he believed q for certain; for knowledge of q might for psychological reasons profoundly alter his whole system of beliefs [Ramsey 1931, p. 180; cf. however, p. 192].

Other reservations about the adequacy of conditionalization as an exclusive model for belief revision center around its assumption about

the form in which new information is received. Indeed, Jeffrey's original philosophical motivation for introducing probability kinematics was his belief that "It is rarely or never that there is a proposition for which the direct effect of an observation is to change the observer's degree of belief in that proposition to 1" (Jeffrey 1968, p. 171). Similar criticisms have been raised by Shafer (1979, 1981), whose theory of belief functions is a more radical attempt to deal with the problem. Both hold that conditioning on an event requires the assignment of an initial probability for that event, prior (in principle at least) to its observation, and for many classes of sensory experiences this seems forced, unrealistic, or impossible.

For example, suppose we are about to hear one of two recordings of Shakespeare on the radio, to be read by either Olivier or Gielgud, but are unsure as to which,and have a prior with mass ½ on Olivier, ½ on Gielgud. After hearing the recording, one might judge it fairly likely, but by no means certain, to be by Olivier. The change in belief takes place by direct recognition of the voice; all the integration of sensory stimuli has already taken place at a subconscious level. To demand a list of objective vocal features which we condition on in order to affect the change would be a logician's parody of a complex psychological process.

Another issue is that our "[subjective] probabilities can change in the light of calculations or of pure thought without any change in the <u>empirical</u> data..." (Good 1977, p. 140). I. J. Good terms such probabilities "evolving" or "dynamic" and has discussed them in a number of papers (cf., e.g., Good 1950, p. 49; 1968; 1977). There are serious difficulties in attempting to model such types of belief revision, particularly if, as noted by Savage (1967, p. 308) and others, the new information is a logical or mathematical consequence of the old. For recent progress in this direction, see Garber (1982), Jeffrey (1982).

It is useful in considering these questions to distinguish between the actual, practical application of Bayes' theorem and its use in modelling successive revision in belief of a hypothetical "rational agent". As a practical matter our new beliefs may bear little relation to our old ones; modelling process of change so general seems elusive. Assuming "temporal coherence" results in a plausible description of belief revision with interesting mathematical consequences (convergence of limiting frequencies, asymptotic normality of posterior distributions, etc.). Jeffrey's rule places fewer restrictions on the hypothesized form of belief revision, yet retains enough structure to permit interesting conclusions to emerge (hence the name "probability kinematics").

Jeffrey's rule was introduced in Jeffrey (1957) and is further discussed in Jeffrey (1965, Chapter 11) and Jeffrey (1968). Isaac Levi (1967; 1970, pp. 147-152) is a vigorous critic of Jeffrey's version of probability kinematics, but has been thoroughly rebutted by Jeffrey (1970, especially at pp. 173-179). Jeffrey's idea was partially anticipated by the Oxford astronomer Donkin (1851, p. 356); cf. Boole (1854, pp. 251-252), Whitworth (1901, pp. 162-169, 181-182), Keynes (1921, pp. 176-177). An independent proposal of Jeffrey's rule appears in Griffeath and Snell (1974). The last few years have seen a sudden upsurge of interest in Jeffrey conditionalization; papers have appeared by May and Harper (1976), Teller (1976), Field (1978), Shafer (1981), Williams (1980), van Fraassen (1980), Garber (1980), Domotor (1980), and Armendt (1980).

2. JEFFREY'S RULE OF CONDITIONING

In this section we develop some of the mathematics connected with Jeffrey's rule of conditioning. Formally: Ω is a countable set, $\underline{P}$ and $\underline{P}^*$ are probability measures on the subsets of Ω, and $\{E_i\}$ is a partition of Ω.

2.1 Bayesian Conditioning

Jeffrey's rule of conditioning is a generalization of ordinary conditioning: given the partition $\{E, E^c\}$, if $P^*(E) = 1$ and $P^*(A) = \Sigma P(A|E_i)\, P^*(E_i)$, then $P^*(A) = P(A|E)$. We therefore begin by investigating when one measure P^* can arise from another measure P by conditioning. To be precise, suppose P and P^* are measures on a countable space Ω. We will say that <u>P^* can be obtained from P by conditioning</u> if there exists a probability space $(\bar{\Omega}, \bar{G}, Q)$, and events $\{E_\omega\}_{\omega\in\Omega}$, $E_\omega \in \bar{G}$ (to be thought of as "$E_\omega = \omega$ occurred"), such that $Q(E_\omega) = P(\omega)$, and an event $E \in \bar{G}$ such that $Q(E) > 0$ and $Q(E_\omega|E) = P^*(\omega)$.

<u>Theorem 2.1</u>: P^* can be obtained from P by conditioning if and only if

(2.1) $\quad P^*(\omega) \leq B\, P(\omega)$ for some constant $B \geq 1$ and all ω.

<u>Proof</u>: If P^* can be obtained from P by conditioning, let $(\bar{\Omega}, \bar{G}, Q)$, $\{E_\omega\}$, E be given. Then for any $\omega \in \Omega$,

$$P^*(\omega) = Q(E_\omega|E) \leq \frac{Q(E_\omega)}{Q(E)} = \frac{P(\omega)}{Q(E)} \quad .$$

This gives (2.1) with $B = 1/Q(E)$.

Conversely, suppose (2.1) is satisfied. Let $\bar{\Omega} = \Omega \times \{a,b\} = \{(\omega,a), (\omega,b)\}\ \omega \in \Omega$. Let $\bar{\mathcal{G}}$ be the set of all subsets of $\bar{\Omega}$. Let $E_\omega = (\omega,a) \cup (\omega,b)$, and let $E = \bigcup_{\omega\in\Omega} (\omega,a)$. Solving the problem of finding Q formally leads to introducing a parameter t, $0 < t < 1$ (t will turn out to be $Q(E)$), and setting

$$Q((\omega,a)) = t\, P^*(\omega)$$

$$Q((\omega,b)) = P(\omega) - t\, P^*(\omega)\ .$$

Because (2.1) is satisfied, t can be chosen small enough so that $Q((\omega,b)) \geq 0$. It is then straightforward to check that Q is a probability on $\bar{\Omega}$ satisfying $Q(E_\omega) = P(\omega)$ and $Q(E_\omega|E) = Q((\omega,a))/\Sigma Q((\omega,a)) = P^*(\omega)$ as required. □

Condition (2.1) places a restriction on P, P^* when both have countable support (but not when both have finite support and $\text{supp}(P^*) \subseteq \text{supp}(P)$). For example, no geometric distribution can be obtained from a Poisson distribution by conditioning, but any Poisson distribution can be obtained from any geometric. If Ω is uncountable, (2.1) can be replaced by the conditions $P^* << P$ and $\frac{dP^*}{dP} \in L_\infty$; see Section 6.

2.2 Jeffrey Conditionalization and Sufficiency

In the example discussed in Section 1, the partition $\{E_i\}$ naturally arose in the course of constructing P^* from P. But one might instead envisage being given another person's $\{P,P^*\}$ and then trying to reconstruct a possible partition $\{E_i\}$ from which the pair $\{P,P^*\}$ could have arisen via Jeffrey conditionalization. Unlike Bayesian conditionalization, this turns out to be always possible.

To apply Jeffrey's rule, it is required to find a partition $\{E_i\}$ such that

$$P(A|E_i) = P^*(A|E_i) \quad \text{for all } A \text{ and } i .$$

This is simply the problem of finding a <u>sufficient partition</u> for the two-element family $\mathfrak{F} = \{P, P^*\}$; see Blackwell and Girshick (1954), Chapter 8. This simple observation makes possible the translation of the ideas of minimal sufficiency and likelihood ratio into the language of Jeffrey's rule.

A partition $\{E_i\}$ is said to be <u>coarser</u> than a second $\{\tilde{E}_j\}$ if every E_i is a union of sets in $\{\tilde{E}_j\}$. For purposes of updating probability, a coarser partition has the advantage that P^* need be specified on fewer sets. A coarsest sufficient partition is said to be minimal sufficient. The following (well known) theorem gives an alternate version of Jeffrey's rule and states that there is always a coarsest partition for which Jeffrey's rule is valid. Some philosophical implications of this fact are discussed by van Fraassen (1980).

<u>Theorem 2.2</u>: Let P, P^* be probability measures with common support on the countable set Ω. If $\{E_i\}$ is a countable partition of Ω such that $P(A|E_i) = P^*(A|E_i)$ for all subsets A and elements of the partition E_i , then for each $\omega \in \Omega$,

$$P^*(\omega) = \frac{P^*(E_i)}{P(E_i)} P(\omega), \quad \omega \in E_i . \tag{2.2}$$

If $R = \{x : P^*(\omega)/P(\omega) = x,\ \omega \in \Omega\}$, and $E_x = \{\omega : P^*(\omega)/P(\omega) = x,\ \omega \in \Omega\}$, then $\{E_x : x \in R\}$ is a minimal sufficient partition for $\{P, P^*\}$.

<u>Proof</u>: The first statement is a version of the Fisher-Neyman factorization theorem; for the second, see Blackwell and Girshick (1957, p.221). □

The following example illustrates the use of the likelihood ratio form of Jeffrey's rule.

Example 2.1 (Whitworth 1901, pp. 167-168):

Question 138. A, B, C were entered for a race, and their respective chances of winning were estimated at $\frac{2}{11}, \frac{4}{11}, \frac{5}{11}$. But circumstances come to our knowledge in favour of A, which raise his chance to $\frac{1}{2}$; what are now the chances in favour of B and C respectively?

Answer. A could lose in two ways, viz. either by B winning or by C winning, and the respective chances of his losing in these ways were *à priori* $\frac{4}{11}$ and $\frac{5}{11}$, and the chance of his losing at all was $\frac{9}{11}$. But after our accession of knowledge the chance of his losing at all becomes $\frac{1}{2}$, that is, it becomes diminished in the ratio of 18 : 11. Hence the chance of either way in which he might lose is diminished in the same ratio. Therefore the chance of B winning is now

$$\frac{4}{11} \times \frac{11}{18}, \text{ or } \frac{4}{18};$$

and of C winning

$$\frac{5}{11} \times \frac{11}{18}, \text{ or } \frac{5}{18}.$$

These are therefore the required chances.

2.3 Generalization

There is a version of Jeffrey's rule which takes the support of the measures P and P^* into account. Call a point $\omega \in \Omega$ a support point of P if $P(\omega) > 0$. Let $\mathrm{supp}(P)$ denote the set of support points of P. In general, P and P^* will not have the same support - indeed with standard conditioning $\mathrm{supp}(P^*)$ is strictly smaller than $\mathrm{supp}(P)$. Clearly P^* will simply have to be freshly quantified on $\mathrm{supp}(P^*) - \mathrm{supp}(P)$. This leads to the following generalized form of Jeffrey's rule:

(2.3) Suppose $\{E_i\}$ is a partition of $S = \mathrm{supp}(P) \cap \mathrm{supp}(P^*)$ such that

(J) $P(\omega|E_i) = P^*(\omega|E_i)$ for all $\omega \in S$ and all i.

Then for any set A,

$$P^*(A) = \Sigma P(A|E_i)\, P^*(E_i) + P^*(A \cap (\mathrm{supp}(P^*) - \mathrm{supp}(P))).$$

In what follows, we will assume that $\mathrm{supp}(P^*) = \mathrm{supp}(P)$. Then Jeffrey's rule simplifies to the form $P^*(A) = \Sigma P(A|E_i)\, P^*(E_i)$ as given in (1.2). All the results we prove have straightforward modifications to the general situation (2.3) by restricting attention to $\mathrm{supp}(P^*) \cap \mathrm{supp}(P)$.

3. SUCCESSIVE UPDATING

In the usual applications of subjective probability, information builds up by successive conditioning. In Bayesian conditionalization the order in which new information is incorporated is irrelevant; in Jeffrey conditionalization the situation is more complex.

3.1 The Problem

Consider an initial probability P which is Jeffrey updated to the new probability $P_{\mathcal{E}}$ based on a partition $\{E_i\}_{i=1}^{e}$ and new probabilities $P_{\mathcal{E}}(E_i) = p_i$, $i = 1, 2, \ldots, e$; clearly $P^*(A|E_i) = P_{\mathcal{E}}(A|E_i) = P(A|E_i)$ holds for our new opinion. (P^* denotes our new opinion, however it is obtained: by Bayes' theorem, Jeffrey's rule, complete requantification or whatever. $P_{\mathcal{E}}$ denotes the specific updated probability measure that results from Jeffrey conditionalization.) We then decide to update based on $\{F_j, q_j\}_{j=1}^{f}$, and indicate this order of updating by $P_{\mathcal{E}\mathcal{F}}$. To use Jeffrey's rule at the second stage we must, of course, accept the (J) condition so $P^*_{\mathcal{E}}(A|F_j) = P_{\mathcal{E}}(A|F_j) = P_{\mathcal{E}\mathcal{F}}(A|F_j)$. Clearly the order of updating matters, since the current opinion dominates:

<u>Example 3.1</u>. $\mathcal{E} = \mathcal{F}$, i.e., our belief on the partition $\{E_i\}$ changes first to p_i and then to q_i. The first revision and second revision differ and we currently believe $P^*(E_i) = q_i$.

<u>Example 3.2</u>. Suppose that in a criminal case we are trying to decide which of four defendants, called a, b, c, d, is a thief. We initially think $P(a) = P(b) = P(c) = P(d) = 1/4$. Evidence is then introduced to show that the thief was probably left-handed. The evidence does not demonstrate that the thief was definitely left-handed but leads us to conclude that $P(\text{thief left-handed}) \approx .8$. If a and b are the defendants who are

left-handed, then $E_1 = \{a,b\}$, $E_2 = \{c,d\}$ and $P_{\mathcal{E}}(E_1) = .8$, $P_{\mathcal{E}}(E_2) = .2$. If the <u>only effect</u> of the evidence was to alter the probability of left-handedness - in the sense that $P(A|E_i) = P_{\mathcal{E}}(A|E_i)$ - then $P_{\mathcal{E}}$ is obtained from Jeffrey's rule as $P_{\mathcal{E}}(a) = .4$, $P_{\mathcal{E}}(b) = .4$, $P_{\mathcal{E}}(c) = .1$, $P_{\mathcal{E}}(d) = .1$. Evidence is next presented that it is somewhat likely that the thief was a woman. If the female defendants are a and c, then $F_1 = \{a,c\}$, $F_2 = \{b,d\}$. If $P_{\mathcal{E}\mathcal{F}}(F_1) = .7$ and Jeffrey updating is again judged acceptable, then

$$P_{\mathcal{E}\mathcal{F}}(a) = .56, \quad P_{\mathcal{E}\mathcal{F}}(b) = .24, \quad P_{\mathcal{E}\mathcal{F}}(c) = .14, \quad P_{\mathcal{E}\mathcal{F}}(d) = .06 .$$

If instead the evidence $(F_1, .7)$, $(F_2, .3)$ is presented first and $(E_1, .8)$, $(E_2, .2)$ is presented second, is $P_{\mathcal{F}\mathcal{E}}$ equal to $P_{\mathcal{E}\mathcal{F}}$? Example 3.1 shows that in general the order matters since the currently held opinion governs; in this example the reader may check that order does not matter. We now investigate why.

3.2 Commutativity

There are two aspects to successive updating:

The updating information at each stage:

$$\{E_i, p_i\}_{i=1}^{e}, \quad \{F_j, q_j\}_{j=1}^{f} ; \tag{3.1}$$

the J condition at each stage:

$$P^*(A|E_i) = P(A|E_i) \quad \text{and} \quad P_{\mathcal{E}}^*(A|F_j) = P_{\mathcal{E}}(A|F_j) \tag{3.2}$$

or, if updating is being considered in the other order,

$$P^*(A|F_j) = P(A|F_j) \quad \text{and} \quad P_{\mathcal{F}}^*(A|E_i) = P_{\mathcal{F}}(A|E_i) .$$

The J condition is an internal or psychological condition which must be checked or accepted at each stage. Mathematics has nothing to offer here.

Mathematics can be used to check if (3.1) is compatible with commutativity. Since Jeffrey updating fixes the probabilities on the partition (i.e., $P_{\mathcal{E}\mathcal{F}}(F_j) = q_j$ and $P_{\mathcal{F}\mathcal{E}}(E_i) = p_i$), commutativity will be possible only if

$$(3.3) \qquad P_{\mathcal{E}\mathcal{F}}(E_i) = p_i \quad \text{and} \quad P_{\mathcal{F}\mathcal{E}}(F_j) = q_j \ , \qquad \text{for all } i \text{ and } j \ .$$

It turns out that this condition is sufficient:

Theorem 3.1: If (3.3) holds, then $P_{\mathcal{E}\mathcal{F}} = P_{\mathcal{F}\mathcal{E}}$.

In other words, whenever $P_{\mathcal{F}\mathcal{E}}$ and $P_{\mathcal{E}\mathcal{F}}$ both incorporate (3.1), they actually coincide. Theorem 3.1 is an immediate consequence of Csiszar (1975, Theorem 3.2) and its proof is omitted. Csiszar's theorem implies that the common measure $P_{\mathcal{E}\mathcal{F}} = P_{\mathcal{F}\mathcal{E}}$ is the "I-projection" of the original measure P onto the set of measures which incorporate (3.1). We discuss I-projection further in Section 6.

3.3 Jeffrey Independence

A second approach to the mathematical aspects of commutativity of successive Jeffrey updating uses independence. Two partitions $\mathcal{E} = \{E_i\}$, $\mathcal{F} = \{F_j\}$ such that $P(E_i) > 0$, $P(F_j) > 0$ for all i and j, are P-independent if

$$(3.4) \qquad P(E_i|F_j) = P(E_i) \quad \text{and} \quad P(F_j|E_i) = P(F_j) \ , \quad \text{all } i, j \ .$$

Independence says that conditioning on $\mathcal{F}$ does not change the probabilities on $\mathcal{E}$ and vice versa. Analogously, we define:

(3.5) $\underline{\mathcal{E} \text{ and } \mathcal{F} \text{ are Jeffrey independent}}$ with respect to $P, \{p_i\}$ and $\{q_j\}$ if $P_{\mathcal{E}}(F_j) = P(F_j)$ and $P_{\mathcal{F}}(E_i) = P(E_i)$ holds for all i and j. (Briefly, "J-independent wrt $\{p_i\}$, $\{q_j\}$".) Thus Jeffrey independence says that Jeffrey updating on $\mathcal{E}$ with probability p_i does not change the probability on $\mathcal{F}$ and similarly with $\mathcal{E}$ and $\mathcal{F}$ interchanged. The next theorem shows the connection with commutativity.

<u>Theorem 3.2</u>: Let P, $\{E_i, p_i\}$ and $\{F_j, q_j\}$ be given. Then $P_{\mathcal{E}\mathcal{F}} = P_{\mathcal{F}\mathcal{E}}$ if and only if $\mathcal{E}$ and $\mathcal{F}$ are Jeffrey independent with respect to P, $\{p_i\}$, $\{q_j\}$.

<u>Proof</u>: Note that $P_{\mathcal{E}\mathcal{F}}(A) = P_{\mathcal{F}\mathcal{E}}(A)$ for all events A if and only if

$$(3.6) \qquad \sum_{ij} \frac{p_i q_j}{P_{\mathcal{E}}(F_j)\ P(E_i)} P(AE_iF_j) = \sum_{ij} \frac{p_i q_j}{P_{\mathcal{F}}(E_i)\ P(F_j)} P(AE_iF_j) .$$

Choose $A = E_{i_0} F_{j_0}$ to get

$$P_{\mathcal{E}}(F_{j_0})\ P(E_{i_0}) = P_{\mathcal{F}}(E_{i_0})\ P(F_{j_0}) \quad \text{for all pairs } i_0, j_0 .$$

Keeping i_0 fixed and summing over j_0 yields

$$(3.7a) \qquad P(E_{i_0}) = P_{\mathcal{F}}(E_{i_0}) ;$$

similarly, fixing j_0 and summing over i_0 yields

$$(3.7b) \qquad P_{\mathcal{E}}(F_{j_0}) = P(F_{j_0}) .$$

Thus, $\mathcal{E}$ and $\mathcal{F}$ are Jeffrey independent with respect to P, $\{p_i\}$, $\{q_j\}$. Conversely, if (3.7) holds, then

$$P_{\mathcal{E}}(F_j)\ P(E_i) = P(F_j)\ P(E_i) = P_{\mathcal{F}}(E_i)\ P(F_j)\ .$$

Using this equality shows that (3.6) holds and so $P_{\mathcal{E}\mathcal{F}} = P_{\mathcal{F}\mathcal{E}}$. □

<u>Theorem 3.3</u>: Two partitions $\mathcal{E}$ and $\mathcal{F}$ are P-independent if and only if $\mathcal{E}$ and $\mathcal{F}$ are Jeffrey independent with respect to <u>any</u> update probabilities $\{p_i\}$ and $\{q_j\}$.

<u>Proof</u>: First suppose $\mathcal{E}$ and $\mathcal{F}$ are P-independent. Then

$$(3.8) \qquad P_{\mathcal{E}}(F_j) = \sum_i P(F_j|E_i)p_i = \sum_i P(F_j)p_i = P(F_j)\ .$$

To see the converse, suppose $\mathcal{E}$ and $\mathcal{F}$ are not P-independent. Then there exist E_{i_0} and F_{j_0} such that $P(F_{j_0} E_{i_0}) \neq P(F_{j_0})$. Pick p_{i_0} sufficiently close to 1. Then

$$\sum_i P(F_{j_0} E_i)p_i \neq P(F_{j_0})\ ,$$

and hence (3.8) entails $P_{\mathcal{E}}(F_{j_0}) \neq P(F_{j_0})$. □

<u>Example 3.3</u>. (J-independence $\not\Rightarrow$ P-independence). Suppose $P(\mathcal{E}_i\mathcal{F}_j)$ is given by the following table

	$\mathcal{F}_1$	$\mathcal{F}_2$	$\mathcal{F}_3$	
$\mathcal{E}_1$	1/4	1/8	1/8	1/2
$\mathcal{E}_2$	1/8	0	1/8	1/4
$\mathcal{E}_3$	1/8	1/8	0	1/4
	1/2	1/4	1/4	

Then $\mathcal{E}$ and $\mathcal{F}$ are not independent, but update probabilities $\underset{\sim}{p}$, $\underset{\sim}{q}$ exist such that $\mathcal{E}$ and $\mathcal{F}$ are J-independent with respect to them (see below).

An efficient algorithm for checking J-independence, in this and other examples, is the following. Let r_{ij} denote W. E. Johnson's <u>coefficient of dependence</u> between E_i and F_j (cf. Keynes 1921, pp. 150-155), i.e.,

$$r_{ij} = \frac{P(E_i F_j)}{P(E_i)\ P(F_j)} ,$$

and let $\underset{\sim}{R} = (r_{ij})$; since $\Sigma_i r_{ij} p_i = P_{\mathcal{E}}(F_j)/P(F_j)$ and $\Sigma_j r_{ij} q_j = P_{\mathcal{F}}(E_i)/P(E_i)$, it follows that $\mathcal{E}$ and $\mathcal{F}$ are J-independent with respect to $\{p_i\}$, $\{q_j\}$ if and only if

$$(3.8) \qquad \Sigma_i r_{ij} p_i = 1, \text{ all } j \ ; \ \Sigma_j r_{ij} q_j = 1, \text{ all } i.$$

In Example 3.3

$$R = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{pmatrix}$$

and hence, if

$$\underset{\sim}{p} = \left(p, \frac{1-p}{2}, \frac{1-p}{2}\right), \quad 0 < p < 1 , \quad \text{and} \quad \underset{\sim}{q} = \left(q, \frac{1-q}{2}, \frac{1-q}{2}\right), \quad 0 < q < 1 ,$$

then $\underset{\sim}{p}R = \underset{\sim}{1}$, $R\underset{\sim}{q}^t = \underset{\sim}{1}$; thus $\mathcal{E}$, $\mathcal{F}$ are J-independent with respect to $\underset{\sim}{p}$, $\underset{\sim}{q}$.

<u>Remark</u>. It is not hard to show that if at least one of the two partitions $\mathcal{E}$ and $\mathcal{F}$ has only two elements, then J-independence for <u>some</u> $\underset{\sim}{p}$, $\underset{\sim}{q}$ pair is equivalent to P-independence, and hence to J-independence for <u>all</u> $\underset{\sim}{p}$, $\underset{\sim}{q}$.

Lest the reader think that commutativity always occurs when (3.1) can be incorporated, we conclude this section with an example which has $P_{\mathcal{E}\mathcal{F}}(E_i) = p_i$ (and of course $P_{\mathcal{E}\mathcal{F}}(F_j) = q_i$) but such that $P_{\mathcal{F}\mathcal{E}}(F_j) \neq q_i$.

Example 3.4. Let $\mathcal{E} = \{E,\bar{E}\}$, $\mathcal{F} = \{F,\bar{F}\}$, and define P by

	F	$\bar{F}$	
E	1/8	1/4	3/8
$\bar{E}$	3/8	1/4	5/8
	1/2	1/2	

Suppose $p_1 = p_2 = 1/2$ and $q_1 = 7/15$, $q_2 = 8/15$. Then a simple computation shows that $P_{\mathcal{E}\mathcal{F}}(E) = 1/2 = P_{\mathcal{E}\mathcal{F}}(\bar{E})$, but $P_{\mathcal{F}\mathcal{E}}(F) \neq q_1$.

4. COMBINING SEVERAL BODIES OF EVIDENCE

Suppose we undergo a complex of experiences that result in our simultaneously adopting new degrees of belief on two partitions $\mathcal{E} = \{E_i\}$ and $\mathcal{F} = \{F_j\}$, say

$$P^*(E_i) = p_i \quad \text{and} \quad P^*(F_j) = q_j \ . \tag{4.1}$$

How should we revise our subjective probabilities so as to incorporate these new beliefs? In general, the theory put forth by de Finetti has no neat mathematical answer to this question - you just have to think about things and quantify your opinion as best you can. In this section we discuss two reasonable routes through this quantification procedure. The routes are reasonable in the same sense that exchangeability is a reasonable thing to consider when attempting to quantify probabilities on repeated events - the circumstances which make them subjectively acceptable occur frequently. We first discuss whether measures satisfying (4.1) exist and if so, how to uniquely select one.

4.1 Coherence of P^*

If we are to adopt the degrees of belief P^* in (4.1), they must at least be coherent, i.e., P^* must be extendable to a probability measure. Theorem 4.1 provides a simple necessary and sufficient condition for the existence of such extensions. The proof, given below in the Appendix, gives an efficient algorithm for computing P^* when both partitions are finite.

<u>Theorem 4.1</u>: Let Ω be a countable set, $\mathcal{E} = \{E_i\}$ and $\mathcal{F} = \{F_j\}$ two partitions of Ω, and P, Q two probability measures on $\mathcal{E}$ and $\mathcal{F}$

respectively. There exists a probability measure P^* on Ω such that (4.1) holds if and only if whenever disjoint sets A and B are given, with A a union of elements of $\mathcal{E}$, B a union of elements of $\mathfrak{F}$,

(4.2) $$P(A) + Q(B) \leq 1 .$$

<u>Remark</u>. Condition (4.2) is necessary but not sufficient for Theorem 4.1 to hold if Ω is uncountable.

4.2 <u>Extending P^*</u>

If (4.1) is coherent, it remains to

(4.3) choose a probability P^* on the partition $\{E_i \cap F_j\}$ which agrees with (4.1);

(4.4) extend P^* to all of Ω.

If judged valid, the easiest way of accomplishing (4.3) is to use <u>independence</u>: $P^*(E_i \cap F_j) = P^*(E_i)\, P^*(F_j) = p_i q_j$.

Richard Jeffrey (1957, Chapter 4) has advocated another route from (4.1) to a final probability assignment: successive Jeffrey updating on $\mathcal{E}$ and $\mathfrak{F}$. This raises two issues:

(4.5) When does successive updating satisfy (4.1)?

(4.6) When is successive updating reasonable?

Question (4.5) arises because $P_{\mathcal{E}\mathfrak{F}}$ need not equal $P_{\mathfrak{F}\mathcal{E}}$. Indeed, Example 3.4 provides a situation where (4.1) is coherent because $P_{\mathcal{E}\mathfrak{F}}$ satisfies (4.1), but $P_{\mathcal{E}\mathfrak{F}} \neq P_{\mathfrak{F}\mathcal{E}}$. Since matters are simplified when $P_{\mathcal{E}\mathfrak{F}} = P_{\mathfrak{F}\mathcal{E}}$, we note that the results of Section 3 imply that the following three conditions are equivalent

(4.7a) $P_{\mathcal{E}\mathfrak{F}}(A) = P_{\mathfrak{F}\mathcal{E}}(A)$ for all sets A.

(4.7b) $P_{\mathcal{E}\mathfrak{F}}(E_i) = P_{\mathfrak{F}\mathcal{E}}(E_i)$ and $P_{\mathfrak{F}\mathcal{E}}(F_j) = P_{\mathcal{E}\mathfrak{F}}(F_j)$ for all i and j.

(4.7c) $P_{\mathfrak{F}}(E_i) = P(E_i)$ and $P_{\mathcal{E}}(F_j) = P(F_j)$ for all i and j.

Even when the order does not matter, we still have the responsibility of justifying the resort to successive updating, i.e., problem (4.6). One approach to this is via checking the Jeffrey condition at each stage of updating. This is a somewhat unorthodox mental exercise given we currently believe (4.1), a condition involving both partitions. If we update first on $\mathcal{E}$, then we must check $P(A|E_i) = P^*(A|E_i)$ which amounts to thinking as if we don't know about $\mathfrak{F}$ and are only thinking about $\mathcal{E}$. At the second stage, one then checks $P_{\mathcal{E}}(A|F_j) = P^*_{\mathcal{E}}(A|F_j)$, comparing one's opinion not knowing $\mathfrak{F}$ to one's opinion knowing $\mathfrak{F}$. Examples such as Example 3.4 show that this can be tricky. It is a possible route, however, one more general than the route using independence suggested before.

<u>Remark 1</u>. There is no reason to have $P_{\mathcal{E}\mathfrak{F}} = P_{\mathfrak{F}\mathcal{E}}$ for successive updating to be useful and valid. If each of the (J) conditions is judged valid in forming $P_{\mathcal{E}\mathfrak{F}}$ and if $P_{\mathcal{E}\mathfrak{F}}$ satisfies (4.1), then $P_{\mathcal{E}\mathfrak{F}}$ is a consistent quantification of current belief.

<u>Remark 2</u>. Condition (4.7) implies that $P_{\mathcal{E}\mathfrak{F}}$ and $P_{\mathfrak{F}\mathcal{E}}$ cannot both incorporate (4.1) and both be judged acceptable updates (in the sense that the (J) conditions have been checked) without $P_{\mathcal{E}\mathfrak{F}} = P_{\mathfrak{F}\mathcal{E}}$. Thus non-commutativity is not a real problem for successive Jeffrey updating.

5. MECHANICAL UPDATING

The approach we have taken thus far to justifying Jeffrey's rule is subjective - through checking condition (J). Several authors - Griffeath and Snell (1974), May and Harper (1976), Williams (1980), and van Fraassen (1980) - have pursued a different justification. Given a prior P, partition $\{E_i\}$, and a new measure P^* on $\{E_i\}$, find the "closest" measure to P which agrees with P^* on the partition and take this as defining P^* on the whole space. Since this way of proceeding does not attempt to quantify one's new degrees of belief via introspection, we call this approach mechanical updating.

5.1 Minimum Distance Properties

If "close" is defined in any of several common ways, the closest measure is that given in Jeffrey's rule. We illustrate this with three well known notions of closeness between measures P and Q on the countable set Ω:

(5.1) The variation distance

$$\|P-Q\| = \sup\{|P(B) - Q(B)| : B \subset \Omega\} .$$

Two measures are close in variation distance if they are uniformly close on all subsets.

(5.2) The Hellinger distance

$$H(P,Q) = \sum_{\omega} (\sqrt{P(\omega)} - \sqrt{Q(\omega)})^2 .$$

(5.3) The Kullback-Leibler number of Q with respect to P

$$I(Q,P) = \sum_{\omega} Q(\omega) \log (Q(\omega)/P(\omega)) .$$

The variation and Hellinger distances are actual metrics on the space of probability distributions, the Kullback-Leibler number is not, being asymmetric in its arguments. Kailath (1967) and Csiszár (1977) are good surveys with references of the properties of (5.1), (5.2), and (5.3).

Theorem 5.1: Let Ω be a countable set, P a probability on Ω, and $\{E_i\}$ be a partition of Ω. Suppose $P^*(E_i) \geq 0$ are given numbers such that $\Sigma P^*(E_i) = 1$. Let Q be a probability on Ω such that $Q(E_i) = P^*(E_i)$. Then

$$\|Q-P\| \geq \frac{1}{2} \sup |P(E_i) - P^*(E_i)|, \tag{5.4}$$

$$H(Q,P) \geq \Sigma(\sqrt{P(E_i)} - \sqrt{P^*(E_i)})^2 , \tag{5.5}$$

$$I(Q,P) \geq \Sigma P^*(E_i) \log (P^*(E_i)/P(E_i)) . \tag{5.6}$$

In (5.5), and (5.6) equality holds if and only if $Q(A) = \Sigma P(A|E_i)\, P^*(E_i)$.

Remarks. 1) Although the probability measure given by Jeffrey's rule minimizes the variation distance, it does not do so uniquely; see May (1976). 2) In Theorem 5.1, the minimum distance between P and Q is the distance between P and Q viewed as measures on the partition $\{E_i\}$. 3) A result like Theorem 5.1 holds for several other notions of distance; see Section 6 where a generalization of Theorem 5.1 is given.

5.2 I-Projections and the IPFP

Mechanical updating allows the possibility of updating on more general collections of sets than partitions. Suppose we want to adapt new degrees of belief $P^*(E_i) = p_i$, $1 \leq i \leq n$, where $\mathcal{E} = \{E_1, E_2, \ldots, E_n\}$ is not

necessarily a partition of Ω. This situation is closely related to Jeffrey's proposal of updating simultaneously on several partitions, mentioned in Section 4, in as much as updating simultaneously on partitions $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_k$ is that same as updating on $\mathcal{E} = \bigcup_{i=1}^{k} \mathcal{E}_i$. Conversely, updating on $\mathcal{E} = \{E_1,\ldots,E_n\}$ can be viewed as updating simultaneously on the partitions $\mathcal{E}_1 = \{E_1,E_1^c\}$, $\mathcal{E}_2 = \{E_2,E_2^c\}$, ..., $\mathcal{E}_n = \{E_n,E_n^c\}$. In general, the set $C = \{Q : Q(E_i) = p_i \text{ for all } i\}$ is a convex set of probability measures on Ω which can be empty, contain a single element, or contain many elements. In the first case P^* is incoherent, in the second, P^* is uniquely determined. When the third case holds, we can use the Kullback-Leibler number as a notion of "distance" to pick a unique number of C closest to P.

Theorem 5.2: Let $S(P,\infty) = Q : I(Q,P) < \infty$. If $S(P,\infty) \cap C \neq \phi$, then there exists a unique element $Q_J \in C$ such that $I(Q_J,P) = \inf\{I(Q,P) : Q \in C\}$.

Proof: This is an immediate consequence of Csiszár (1975, Theorem 2.1), C being convex and closed with respect to the variation distance. □

In Csiszár's terminology, Q_J is the I-projection of P onto C. (The term is meant to suggest the projection of a vector in R^n onto a subspace.) The I-projection is closely related to a widely used technique in the statistical analysis of contingency tables.

A standard method of adjusting an $r \times c$ contingency table so that it has given marginal totals is the iterated proportional fitting procedure (IPFP). In this, one first adjusts the table to have specified row sums, say (by dividing the numbers of a given row by the appropriate factor), next adjusts the new table to have the right column sums, and then continues iteratively. It follows from Csiszár (1975, Theorem 3.2) that this procedure converges to the I-projection of the initial table onto the set of

tables with the specified row and column sums. That is, the IPFP finds the "closest" table to the original table with the prescribed margins (provided, of course, this set is nonempty). This is essentially the same as finding the closest measure to an initial probability with prescribed values on two partitions.

The IPFP can be used to compute Q_J of Theorem 5.2 by treating the problem as an n-dimensional contingency table with given margins $P^*(E_1)$, $1 - P^*(E_1)$, ...

5.3 Comparing Different Metrics

Theorem 5.1 suggests that Jeffrey's rule is an uncontroversial form of mechanical updating in the sense that it agrees with virtually every minimum distance rule. As noted above, in the case of two or more partitions, the I-projection or maximum entropy solution can be viewed as a limiting form of successive Jeffrey updating. This is perhaps of some interest inasmuch as mechanical updating via the other minimum distance methods need not, in general, yield the same answer as the I-projection.

Example 5.1. (I-projection $\neq$ minimum variation distance.) Consider passing from an initial table

$$\underset{\sim}{P}^\circ = \begin{array}{|c|c|} \hline 1/4 & 1/4 \\ \hline 1/4 & 1/4 \\ \hline \end{array} \quad \text{to} \quad \underset{\sim}{P} = \begin{array}{c|c|c|} & 1/3 & 2/3 \\ \hline 1/3 & p_1 & p_2 \\ \hline 2/3 & p_3 & p_4 \\ \hline \end{array},$$

a new table with the specified margins and which is otherwise as "close" to the original table as possible, according to some notion of closeness.

a) The table P^I given by $p_1 = \frac{1}{9}$, $p_2 = p_3 = \frac{2}{9}$, $p_4 = \frac{4}{9}$ minimizes $I(\underset{\sim}{P},\underset{\sim}{P}^\circ)$ since P° is independent and I-projections preserve the association factor of a 2×2 table (see, e.g., Mosteller (1968), p. 3). The

variation distance for this table is $\|\underset{\sim}{P}^I - \underset{\sim}{P}^\circ\| = \frac{1}{2}\sum_{i=1}^{4} |p_i - \frac{1}{4}| = \frac{1}{2}|\frac{1}{9} - \frac{1}{4}| +$ $|\frac{2}{9} - \frac{1}{4}| + \frac{1}{2}|\frac{4}{9} - \frac{1}{4}| = \frac{7}{36}$.

b) To find the table $\underset{\sim}{P}^V$ with minimum variation distance from $\underset{\sim}{P}^\circ$, subject to the margin constraints, note that given p_1, one has $p_2 = p_3 = \frac{1}{3} - p_1$ and $p_4 = p_1 + \frac{1}{3}$. Hence

$$\|\underset{\sim}{P} - \underset{\sim}{P}^\circ\| = \frac{1}{2}\Sigma|p_i - \frac{1}{4}| = \frac{1}{2}\left\{|\frac{1}{4} - p_1| + 2|\frac{1}{12} - p_1| + |-\frac{1}{12} - p_1|\right\}$$

which is minimized by $p_1 = \frac{1}{12}$, the median of $\{-\frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{4}\}$. Hence $\underset{\sim}{P}^V = (\frac{1}{12}, \frac{1}{4}, \frac{1}{4}, \frac{5}{12})$ and $\|\underset{\sim}{P}^V - \underset{\sim}{P}^\circ\| = \frac{1}{6}$. □

There has been considerable interest recently in maximum entropy methods, especially in the philosophical literature (Rosenkrantz (1979), Williams (1980), van Fraassen (1980)). Example 5.1 suggests that any claims to the effect that maximum entropy revision is the only correct route to probability revision should be viewed with considerable caution because of its strong dependence on the measure of closeness being used.

6. ABSTRACT PROBABILITY KINEMATICS

In this section we briefly discuss the generalization of Jeffrey's rule of conditioning from the countable setting to general spaces. The need for such a generalization is shown by passing to the limit in the example of Section 1.1.

Example 6.1. Consider an infinite sequence of zero or one outcomes $X_1, X_2, X_3, \ldots$. Suppose that the joint distribution of X_i is exchangeable and set $S_n = X_1 + \ldots + X_n$. Then, as shown by de Finetti, the limit

$$Z = \lim_{n\to\infty} \frac{S_n}{n}$$

exists almost surely and

$$P(S_n = k | Z = p) = \binom{n}{k} p^k (1-p)^{n-k} .$$

One consequence of de Finetti's theorem is that one may decide on a subjective probability distribution for an infinite exchangeable sequence of coin-tosses by introspecting on the "prior distribution" $P\{Z \in dp\} \equiv d\mu(p)$. In the example of Section 1.1, the effect of the informant's information could be taken into account by choosing a new prior $d\mu^*(p)$ and Jeffrey's rule becomes:

$$P^*(S_n = k) = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} \, d\mu^*(p) .$$

This illustrates the use of Jeffrey updating via a continuous "sufficient statistic" rather than a countable "sufficient partition". The generalization we use replaces partitions by σ-algebras.

Consider a probability space $(\Omega, \mathcal{G}, P)$, thought of as describing our current subjective beliefs about the σ-algebra of events $\mathcal{G}$. Let P^* be a new probability measure on $\mathcal{G}$ and $\mathcal{G}_0 \subseteq \mathcal{G}$ a sub-σ-algebra of $\mathcal{G}$. Let C be an $\mathcal{G}_0$-measurable set such that $P(C) = 0$ and $\bar{P} << \bar{P}^*$ on $\Omega - C$, where $\bar{P}, \bar{P}^*$ are the restrictions of P, P^* to $\mathcal{G}_0$. The appropriate version of Jeffrey's condition (J) is:

(J') $\qquad \mathcal{G}_0$ is sufficient for $\{P, P^*\}$.

When condition (J') holds, Jeffrey's rule of conditioning becomes:

$$P^*(A) = \int_{\Omega-C} P(A|\mathcal{G}_0)\, P^*(d\omega) + P^*(A \cap C) , \tag{6.1}$$

where $P(A|\mathcal{G}_0)$ is the conditional probability of A given $\mathcal{G}_0$. If $P^* << P$, we can take $C = \phi$.

Much of the mathematical machinery for dealing with Jeffrey conditionalization in this generality has been developed (for a different purpose) by Csiszár (1967). His Lemma 2.2 translates into a likelihood ratio version of Jeffrey's rule (compare (2.2)): Let λ be a σ-finite measure which dominates P, P^*. Let $\bar{\lambda}, \bar{P}\ \bar{P}^*$ be the restrictions to $\mathcal{G}_0$. Assume $\bar{\lambda}$ is σ-finite. Let $\bar{p}(x)\ \bar{p}^*(x)$ be the densities of P, P^* with respect to $\bar{\lambda}$ and p^* the density of p^* with respect to λ. If condition (J') holds, then:

$$p^*(x) = \begin{cases} \bar{p}^*(x)/\bar{p}(x) & \text{if } \bar{p}(x) > 0 \\ p^*(x) & \text{if } \bar{p}(x) = 0 . \end{cases} \tag{6.2}$$

Identity (6.2) is a version of the Fisher-Neyman factorization theorem (see Halmos and Savage (1949)).

Csiszár's results allow us to give a single theorem which includes Theorem 5.1, showing that the closest measure to P which agrees with P^* on $\mathcal{G}_0$ is the measure given by (6.1). Csiszár has introduced the notion of f-divergence, where f is a convex function defined in the interval $(0,\infty)$. If μ_1 and μ_2 are two measures on $(\Omega, \mathcal{G})$, the f-divergence of μ_1 and μ_2 is

$$I_f(\mu_1,\mu_2) = \int p_2(x)\, f\left(\frac{p_1(x)}{p_2(x)}\right) \lambda(dx) ,$$

where $\mu_i << \lambda$ and $p_i = \frac{d\mu_i}{d\lambda}$ $(i = 1,2)$. Taking $f(u) = u \log u$ gives the Kullback-Leibler number, $f(u) = (u^{1/2} - 1)^2$ the Hellinger distance, $f(u) = |u - 1|$ the variation distance. Csiszár shows that several other notions of distance are also f-divergences for an appropriate f.

Theorem 6.1: Let C be the set of probability measures on $(\Omega, \mathcal{G})$ which agree with P^* on $\mathcal{G}_0$, and f a convex function on $(0,\infty)$. Then under condition (J'),

$$(6.3) \qquad I_f(P^*,P) = I_f(\bar{P}^*,\bar{P}) = \inf\{I_f(Q,P) : Q \in C\} .$$

If f is strictly convex, then P^* is the unique probability measure on $\mathcal{G}$ which minimizes the right-hand side of (6.3).

Proof: The first equality follows from the sufficiency of $\mathcal{G}_0$ for $\{P,P^*\}$, the second from Csiszár's (1967, Section 3) version of the minimum information discrimination theorem of Kullback and Leibler: $I_f(Q,P) \geq I_f(\bar{Q},\bar{P})$. Since $I_f(\bar{Q},\bar{P}) = I_f(\bar{P}^*,\bar{P})$, (6.3) follows. If f is strictly convex, $I_f(\cdot,P)$ is also, and the theorem follows. $\square$

APPENDIX

We first prove a slight generalization of Theorem 4.1:

Theorem 4.2: Let Ω be a countable set, S a σ-algebra of subsets, $\mathcal{A}$ and $\mathcal{B}$ sub-σ-algebras of S, and μ and ν probability measures on $\mathcal{A}$ and $\mathcal{B}$ respectively. A necessary and sufficient condition for there to exist a probability measure P on (Ω,S) such that P equals μ on $\mathcal{A}$ and P equals ν on $\mathcal{B}$ is

(4.8) for each $A \in \mathcal{A}$ and $B \in \mathcal{B}$ such that $A \cap B = \phi$,

$$\mu(A) + \nu(B) \leq 1 .$$

Proof: The condition is clearly necessary. To prove sufficiency, let $\{A_i\}_{i=1}^{\infty}$ be the atoms of $\mathcal{A}$ and $\{B_i\}_{i=1}^{\infty}$ be the atoms of $\mathcal{B}$. Let $\Omega_a = \{A_i\}_{i=1}^{\infty}$ and $\Omega_b = \{B_i\}_{i=1}^{\infty}$ both thought of as discrete topological spaces. In $\Omega_a \times \Omega_b$, consider the set $F = \bigcup_{A_i \cap B_j \neq \phi} A_i \times B_j$. This is a closed set in $\Omega_a \times \Omega_b$ and according to Theorem 11 of Strassen (1964) a necessary and sufficient condition for there to exist a probability measure γ on F such that γ has margins μ and ν is that for every $B \in \mathcal{B}$,

(4.9) $$\nu(B) \leq \mu(\pi_a(F \cap S \times B))$$

where π_a is the projection of a set into its first coordinate. Clearly $\pi_a(F \cap S \times B) = \sum_{A_i \cap B = \phi} A_i$ is the smallest $\mathcal{A}$ measurable set containing B. Thus Strassen's condition (4.9) is satisfied if and only if

(4.10) whenever $A \in \mathcal{A}$, $B \in \mathcal{B}$ and $B \subset A$, $\nu(B) \leq \mu(A)$.

Condition (4.10) is equivalent to (4.8). Hence Strassen's theorem gives a measure γ which may be regarded as a measure on the partition $\{A_i \cap B_j\}$.

Since Ω is countable, γ can clearly be extended to a measure on all of Ω and then restricted to a measure P on S with the desired properties. $\square$

In the proof of Theorem 4.2, we have used Strassen's theorem, which itself uses the Hahn-Banach theorem. When the two partitions are both composed of a finite number of sets, Hansel and Troallic (1978) have shown that Strassen's theorem follows from the max flow-min cut theorem. There are efficient algorithms for finding maximum flows, and hence for checking (4.2), in Bondy and Murty (1976), Chapter 11.

Finally, we consider the extension of Theorem 4.2 to more general spaces.

Let (Ω,S) be a measurable space, let $\mathcal{A}$, $\mathcal{B}$ be sub-σ-algebras of S and let μ and ν be probability measures on $\mathcal{A}$ and $\mathcal{B}$ respectively. When does there exist a probability measure P on $\sigma(\mathcal{A}, \mathcal{B})$ such that P restricts to μ on $\mathcal{A}$ and ν on $\mathcal{B}$? We have argued above that if Ω is <u>countable</u>, then a necessary and sufficient condition is

(4.11) $$\forall A \in \mathcal{A},\ B \in \mathcal{B},\ A \cap B = \phi \Rightarrow \mu(A) + \nu(B) \leq 1 .$$

It is easy to show that (4.11) is in fact necessary and sufficient for the existence of a <u>finitely additive</u> measure $\hat{P}$ on the algebra generated by $\mathcal{A}$ and $\mathcal{B}$ (and hence on the algebra of all subsets of Ω) which restricts to μ on $\mathcal{A}$ and ν on $\mathcal{B}$, even if $\mathcal{A}$ and $\mathcal{B}$ are merely algebras. Briefly, one considers the following linear subspace of bounded real valued functions from Ω: $L = \{f + g:\ f$ is $\mathcal{A}$ measurable and g is $\mathcal{B}$ measurable$\}$, and extends the positive linear functional $\ell(f+g) = \mu(f) + \nu(g)$ using the Hahn-Banach theorem. Condition 4.11 is then used to show ℓ is well defined.

We now present an example due to David Freedman and Jim Pitman which shows that condition (4.11) is not sufficient to ensure that a <u>countably</u> additive extension of μ and ν exists. The example shows a bit more: it shows that Theorem 11 of Strassen (1965) cannot be extended to give conditions for a measure on an F_σ set of the unit square to have given margins.

<u>Example 4.2</u>. (D. Freedman, J. Pitman): There exists an F set K in the unit square and a finitely additive probability π on K which has marginal projections equal to Lebesgue measure on each coordinate but such that K supports no countably additive probability with these margins.

<u>Remark</u>. Taking $\Omega = K$

$$\mathcal{G} = \{(A \times [0,1]) \cap K: \ A \text{ is a Borel set of } [0,1]\} ,$$

$$\mathcal{B} = \{([0,1] \times B \cap K: \ B \text{ is a Borel set of } [0,1]\} ,$$

with μ and ν as Lebesgue measure on $\mathcal{G}$ and $\mathcal{B}$ respectively, gives a situation in which 4.11 is satisfied (since a finitely additive refinement exists) but no countably additive refinement exists.

<u>Proof</u>: To construct K, let a_n be a sequence of numbers in $(0,1)$ with $a_n \uparrow 1$. Let ℓ_n be the line in the unit square connecting $(0,0)$ to $(1,a_n)$. Let $K = \cup_n \ell_n$. Note that K does not include the diagonal. To construct π, let π_n be Lebesgue measure on the Borel sets of ℓ_n. Let ρ be any finitely additive probability measure defined on all subsets of the integers $\{1,2,3,\ldots\}$ such that ρ is zero on finite subsets. Let $\pi(s) = \int \pi_n(s)\, \rho(dn)$. Each π_n, considered as a probability on K, projects to Lebesgue measure on the x-axis of the unit square. Further, the projection of π_n onto the y-axis of the unit square gives Lebesgue

measure restricted to the set $\{(0,y): 0 \leq y \leq a_n\}$. It follows easily that the y-axis margin of is Lebesgue measure. It only remains to argue that K does not support a countably additive probability measure P which projects to Lebesgue measure. If $X: K \to [0,1]$, and $Y: K \to [0,1]$ are the two projections, then $E(X) = E(Y) = \frac{1}{2}$ because each of X and Y are uniformly distributed by construction. Any countably additive P would have to put positive probability on some line ℓ_n and since for all $(x,y) \in K$, $y < x$, this forces $E(X) > E(Y)$. The contradiction shows P cannot exist. $\square$

Persi Diaconis is Professor, Department of Statistics, Stanford University, Stanford, California 94305. Sandy L. Zabell is Associate Professor, Department of Mathematics, Northwestern University, Evanston, Illinois 60201. The authors are grateful to I. J. Good, Richard C. Jeffrey, Sherry May, and Amos Tversky for helpful comments and suggestions.

REFERENCES

Armendt, B. (1980). "Is there a Dutch book argument for probability kinematics?", Philosophy of Science 47, 583-588.

Blackwell, D., and Girshick, M. A. (1954). Theory of Games and Statistical Decisions, New York: John Wiley and Sons.

Bondy, J. A., and Murty, U. S. R. (1976). Graph Theory with Applications, New York: American Elsevier.

Boole, G. (1854). An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities, London: Macmillan.

Csiszár, I. (1967). "Information type measures of difference of probability distributions and indirect observations", Studia Sci. Math. Hung. 2, 299-318.

Csiszár, I. (1975). "I-divergence geometry of probability distributions of minimization problems", Annals of Probability 3, 146-158.

Csiszár, I. (1977). "Information measures: A critical survey", Trans. Seventh Prague Conf., (Prague 1974), Prague: Academia, 73-86.

Domotor, Z. (1980). "Probability kinematics and representation of belief change", Philosophy of Science 47, 384-403.

Domotor, Z., Zanotti, M., and Graves, H. (1980). "Probability kinematics", Synthese 44, 421-442.

Donkin, W. F. (1851). "On certain questions relating to the theory of probabilities", Phil. Mag. Series 4, 1, 353-368.

Field, H. (1978). "A note on Jeffrey conditionalization", Philosophy of Science 45, 361-367.

de Finetti, B. (1972). Probability, Induction and Statistics, New York: John Wiley and Sons.

de Finetti, B. (1974). Theory of Probability (Vol. 1), New York: John Wiley and Sons.

de Finetti, B. (1975). Theory of Probability (Vol. 2), New York: John Wiley and Sons.

Fischoff, B. (1975). "Hindsight ≠ foresight: The effect of outcome knowledge on judgement under uncertainty", Journal of Experimental Psychology: Human Perception and Performance 1, 288-299.

Fischoff, B., and Beyth, R. (1975). "I knew it would happen - remembered probabilities of once future things", Organizational Behavior and Human Performance 13, 1-16.

Freedman, D., and Purves, R. (1969). "Bayes method for bookies", Annals of Mathematical Statistics 40, 1177-1186.

Garber, D. (1980). "Discussion: Field and Jeffrey conditionalization", Philosophy of Science 47, 142-145.

Garber, D. (1982). "Old evidence and logical omniscience in Bayesian confirmation theory", Minnesota Studies in Philosophy of Science, forthcoming.

Good, I. J. (1950). Probability and the Weighing of Evidence, New York: Hafner.

Good, I. J. (1968). "Corroboration, explanation, evolving probability, simplicity, and a sharpened razor", British Journal for the Philosophy of Science 19, 123-143.

Good, I. J. (1977). "Dynamic probability, computer chess, and the measurement of knowledge", in Machine Intelligence 8, ed. E. W. Elcock and D. Mitchie, Ellis Horwood Ltd. and John Wylie, 139-150.

Griffeath, D., and Snell, L. J. (1974). "Optimal stopping in the stock market", Annals of Probability 2, 1-13.

Hacking, I. (1967). "Slightly more realistic personal probability", Philosophy of Science 34, 311-325.

Halmos, P., and Savage, L. J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics, Annals of Mathematical Statistics 20, 225-241.

Hansel, G., and Troallic, J. P. (1978). "Measures marginales et théoreme de Ford-Fulkerson", Z. Wahrscheinlichkeitstheorie verw. Gebiete 43, 245-251.

Jamieson, B. (1974). "A Martin boundary interpretation of the maximum entropy argument", Z. Wahrscheinlichkeitstheorie verw. Gebiete 30, 265-272.

Jeffrey, R. (1957). Contributions to the Theory of Inductive Probability, Ph.D. thesis, Princeton University.

Jeffrey, R. (1965). The Logic of Decision, New York: McGraw-Hill.

Jeffrey, R. (1968). "Probable knowledge", in The Problem of Inductive Logic, ed. I. Lakatos, Amsterdam: North-Holland, 166-180.

Jeffrey, R. (1970). "Dracula meets Wolfman: Acceptance vs. partial belief", Induction Acceptance and Rational Belief, ed. M. Swain, Dordrecht: Reidel, 157-185.

Jeffrey, R. (1982). "Bayesianism with a human face", Minnesota Studies in Philosophy of Science, forthcoming.

Kailath, T. (1967). "The divergence and Bhattacharya distance measures in signal detection", IEEE Transactions on Comm. Tech. **15**, 52-60.

Keynes, J. M. (1921). A Treatise on Probability, London: Macmillan.

Kullback, S. (1959). Information Theory and Statistics, New York: John Wiley and Sons.

Lehmann, E. (1959). Testing Statistical Hypotheses, New York: John Wiley and Sons.

Levi, I. (1967). "Probability kinematics", Philosophy of Science **18**, 197-209.

Levi, I. (1970). "Probability and evidence" in Induction, Acceptance, and Rational Belief, ed. Swain, Dordrecht: Reidel, 134-156.

May, S. (1976). "Probability kinematics: a constrained optimization problem", Journal of Philosophical Logic **5**, 395-398.

May, S., and Harper, W. (1976). "Toward an optimization procedure for applying minimum change principles in probability kinematics", in Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science (Vol. 1), eds. W. L. Harper and C. A. Hooker, Dordrecht: Reidel, 137-166.

Mosteller, F. (1968). "Association and estimation in contingency tables", Journal of the American Statistical Association **63**, 1-28.

Ramsey, F. P. (1931). "Truth and probability", in The Foundations of Mathematics and Other Logical Essays, ed. R. B. Braithwaite, London: Routledge and Kegan, P., 156-198. Reprinted in (1964) in Studies in Subjective Probability, eds. H. E. Kyburg, Jr., and H. E. Smokler, New York: John Wiley and Sons, 61-92.

Rosenkrantz, R. D. (1977). Inference, Method, and Decision, Dordrecht: Reidel.

Savage, L. J. (1967). "Difficulties in the theory of personal probability", Philosophy of Science **34**, 305-310.

Shafer, G. (1976). A Mathematical Theory of Evidence, Princeton, New Jersey: Princeton University Press.

Shafer, G. (1979). "Two theories of probability", in PSA 1978, Vol. 2, East Lansing, Michigan: Philosophy of Science Association.

Shafer, G. (1981). "Jeffrey's rule of conditioning", Philosophy of Science 48, 337-362.

Slovic, P., and Fischoff, B. (1977). "On the psychology of experimental surprises", Journal of Experimental Psychology: Human Perception and Performance 3, 544-551.

Strassen, V. (1965). "The existence of probability measures with given marginals", Annals of Mathematical Statistics 36, 423-439.

Teller, P. (1976). "Conditionalization, observation, and change of preference", in Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science (Vol. 1), eds. W. L. Harper and C. A. Hooker, Dordrecht: Reidel, 205-253.

van Fraassen, Bas. C. (1980). "Rational belief and probability kinematics", Philosophy of Science 47, 165-187.

Whitworth, W. A. (1901). Choice and Chance, 5th ed., Cambridge: Deighton Bell and Co.

Williams, P. M. (1980). "Bayesian conditionalization and the principle of minimum information", British Journal for the Philosophy of Science 31, 131-144.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 315	2. GOVT ACCESSION NO. AD-A115397
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) UPDATING SUBJECTIVE PROBABILITY	5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) PERSI DIACONIS and SANDY L. ZABELL	8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0475
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office Of Naval Research Statistics & Probability Program Code 436 Arlington, VA 22217	12. REPORT DATE MARCH 2, 1982
	13. NUMBER OF PAGES 36
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This is a revised version of Technical Report No. 136, October 1979, issued under National Science Foundation Grant MCS77-16974.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Subjective probability; Non-conditioning methods; Exchangeability.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

PLEASE SEE REVERSE SIDE.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

\# 315

UPDATING SUBJECTIVE PROBABILITY

Jeffrey's rule for revising a probability P to a new probability P^* based on a partition $\{E_i\}_{i=1}^n$ is

$$P^*(A) = \Sigma\, P(A|E_i)\, P^*(E_i) \;.$$

This is valid if it is judged that $P^*(A|E_i) = P(A|E_i)$ for all A and i. This paper discusses some of the mathematical properties of ~~this rule~~, connecting it with sufficient partitions, and maximum entropy updating of contingency tables. The main results concern simultaneous revision on two partitions.

FILME

7 – 8